ARTICLE

Risk of Breast Cancer After False-Positive Test Results in Screening Mammography

My von Euler-Chelpin, Louise Madeleine Risør, Brian Larsen Thorsted, Ilse Vejborg

Manuscript received August 30, 2011; revised February 27, 2012; accepted February 28, 2012.

Correspondence to: My von Euler-Chelpin, PhD, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5 opg. B, PO Box 2099, 1014 Copenhagen K, Denmark (e-mail: myeu@sund.ku.dk).

Background   Screening for disease in healthy people inevitably leads to some false-positive tests in disease-free individuals. Normally, women with false-positive screening tests for breast cancer are referred back to routine screening. However, the long-term outcome for women with false-positive tests is unknown.

Methods   We used data from a long-standing population-based screening mammography program in Copenhagen, Denmark, to determine the long-term risk of breast cancer in women with false-positive tests. The age-adjusted relative risk (RR) of breast cancer for women with a false-positive test compared with women with only negative tests was estimated with Poisson regression, adjusted for age, and stratified by screening round and technology period. All statistical tests were two-sided.

Results   A total of 58 003 women, aged 50–69 years, were included in the analysis. Women with negative tests had an absolute cancer rate of 339/100 000 person-years at risk, whereas women with a false-positive test had an absolute rate of 583/100 000 person-years at risk. The adjusted relative risk of breast cancer after a false-positive test was 1.67 (95% confidence interval [CI] 1.45 to 1.88). The relative risk remained statistically significantly increased 6 or more years after the false-positive test, with point estimates varying between 1.58 and 2.30. When stratified by assessment technology phase and using equal follow-up time, the false-positive group from the mid 1990s had a statistically significantly higher risk of breast cancer (RR = 1.65, 95% CI = 1.22 to 2.24) than the group with negative tests, whereas the false-positive group from the early 2000s was not statistically significantly different from the group testing negative.

Conclusions   The implementation of new assessment technology coincided with a decrease in the size of excess risk of breast cancer for women with false-positive screening results. However, it may be beneficial to actively encourage women with false-positive tests to continue to attend regular screening.

J Natl Cancer Inst 2012;104:682–689

In screening mammography, women with positive screening tests typically undergo assessment with triple diagnostics, that is, additional mammograms, ultrasound, palpation, and, if needed, fine needle aspiration cytology or core biopsy. In the majority of instances, the suspicion of malignancy can be ruled out or the final diagnosis of breast cancer can be determined on the basis of these triple diagnostics. In a minority of instances, a surgical biopsy may be needed to reach a conclusion. A high proportion of false-positive tests may result from a wish to uphold high sensitivity, erratic program adherence, technical insufficiency, inadequate interpretive skills, or may lie within the characteristics of the screening population, such as prior benign breast lesions (1).

Women with false-positive tests manifest suspicious mammographic patterns in their breast tissue including tumor-like masses, suspicious microcalcifications, skin thickening or retraction, recently retracted nipples, distortions, asymmetric densities, or suspicious axillary lymph nodes (1). One might therefore hypothesize that these women, despite the thorough assessment procedure to exclude malignancies, are at a higher risk of breast cancer than women without these suspicious patterns in their breast tissue. This hypothesis is supported by the overwhelming evidence for an increased risk of breast cancer in women with benign breast lesions (2–7). To our knowledge, only three short-term studies have followed the breast cancer risk in women with false-positive screening tests, two from the Netherlands (8,9) and one from the United Kingdom (10). In a small Netherlands study (8) from 1988, women with false-positive screening tests had an excess breast cancer risk in the 5 years following screening, whereas the other Netherlands study (9) from 2001 found no excess risk. In the East Anglian screening program, women with false-positive tests had a higher interval cancer rate and a higher detection rate at the subsequent screen than women with negative screening tests (8,10). However, the short-term excess risk identified in the 1988 Netherlands study (8) could largely be attributed to misclassification, and the long-term breast cancer risk in women with false-positive tests remains unknown.

Normally, women with false-positive screening tests are referred back to routine screening. To evaluate whether or not this can be considered a safe policy, or whether a closer follow-up should be considered, it is important to know the long-term fate of women with false-positive tests. We used data from a long-standing population-based screening mammography program to determine the long-term risk of breast cancer in women with false-positive screening tests.

Methods

Screening Setting

Population-based screening mammography started in Copenhagen, Denmark, on April 4, 1991. The screening program was organized in approximately biennial invitation rounds April 4, 1991, to April 23, 1993; April 26, 1993, to May 30, 1995; June 1, 1995, to March 24, 1997; March 25, 1997, to April 19, 1999; April 20, 1999, to March 31, 2001; April 1, 2001, to May 31, 2003; June 1, 2003, to December 31, 2005. In each invitation round, all women aged 50–69 years were personally invited to screening. At first screen, two projections of each breast were made, a craniocaudal and an oblique. At subsequent screens, one projection was made for women with fatty breast tissue and two projections for women with mixed/dense tissue. From 2001 onward, this policy was changed gradually, and from 2004 onward, all women had two projections. At subsequent screens, earlier mammograms were retrieved for comparison. Screen-film mammography was used throughout the study period from 1991 to 2005, and mammograms were evaluated independently by two radiologists, both trained to secure optimal accuracy. In the event of suspicious findings, women were recalled for assessment using the triple test consisting of clinical examination, mammography, and needle biopsy of all palpable solid lesions and of every uncertain, suspicious, or clearly malignant occurrence of disease. From 1992 onward, mammography was supplemented with whole-breast ultrasound examination of palpable and/or mammographically uncertain, suspicious, or malignant lesions. From 1992 onward, ultrasound-guided fine needle aspiration cytology and/or histological biopsy were used in the assessments. High-frequency ultrasound devices were introduced in 2001. Since 2002, suspicious microcalcifications and impalpable mammographic findings that could not be found by ultrasound were examined using stereotactic biopsy equipment. Digital mammography was introduced in 2006, after the end of ascertainment of screening data for this study. Women cleared of suspicion for breast cancer were referred back to routine screening, and women with findings consistent with breast cancer were referred for treatment. In the event of inconsistent findings in the triple test, further investigations were undertaken. If consensus still could not be reached, the women were referred to surgical biopsy (11).

Data

Data on screening results from 1991 to 2005 were retrieved from the Copenhagen Mammography Register containing information on invitation date, participation, and test results. Cancer data were supplied by the Danish Cancer Registry and the Danish Breast Cancer Cooperative Group. The study included breast cancer (C50) and carcinoma in situ (D05) according to the *International*

CONTEXTS AND CAVEATS

Prior knowledge
Women with false-positive tests after mammography screening are generally referred back for routine screening. However, it is not known whether these women have a higher long-term risk for breast cancer compared with women who test negative.

Study design
Data from a population-based mammography screening program from 1991 to 2005 in Denmark was used to assess the risk of breast cancer and ductal carcinoma in situ in women aged 50–69 years who received false-positive test results.

Contribution
Women with negative tests had an absolute cancer rate of 339/100 000 person-years at risk, whereas women with a false-positive test had an absolute rate of 583/100 000 person-years at risk. The relative risk of breast cancer in women with false-positive tests remained statistically significantly higher than in women with negative tests for 6 or more years after the false positive test but decreased after 2000 with newer screening technology.

Implication
Even with newer screening methods, women with false-positive tests should be encouraged to use regular mammographic screening because a false-positive test may indicate underlying pathology that could result in breast cancer.

Limitations
False-positive tests could not be classified by right or left breast because those data were not included in the dataset. The sample sizes and follow-up times for the analysis by technology period were smaller than those in other analyses.

From the Editors

Classification of Disease, *Tenth Revision* (*ICD-10*). By far, the majority of the carcinoma in situ occurrences were ductal carcinoma in situ (DCIS). Linkage on the individual level was achieved with the Danish Civil Registration System number, which is a unique personal identification number. Women with a breast cancer diagnosis before first invitation were excluded from the analysis. Permission for data analysis was granted by the Danish Data Inspection Agency (No. 2008-41-21).

Statistical Analysis

The cancer detection rate for a given invitation round was calculated as number of breast cancers, invasive and DCIS, detected at screening, divided by number of screened women (12,13). In Denmark, all screening mammograms requiring diagnostic assessment are referred to as positive tests without further specification, equivalent to the Breast Imaging Reporting and Data System (BI-RADS)-0 (14). The recall rate for a given invitation round was calculated as the number of women recalled for assessment divided by the number of screened women. A false-positive screening test was defined as any screening test requiring further diagnostic assessment in which neither invasive breast cancer nor DCIS was diagnosed. Women cleared of suspicion of breast cancer at the triple test were referred to as false-positive type 1, and women cleared of suspicion at surgical biopsy were referred to as false-positive type 2.

The false-positive rate for a given invitation round was calculated as number of women with a false-positive test divided by number of screened women.

The relative risk (RR) and the 95% confidence interval (CI) of breast cancer for women with false-positive tests as compared with women with negative tests were estimated with Poisson regression. We undertook three different analyses. First, the incidence rate of breast cancer was analyzed as a log-linear function of attained age (a) and exposure status (s) and expressed as $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where α is the intercept and β is the slope of the regression line. Age was divided into 5-year age groups (50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and 85–89 years), and exposure status was divided into false-positive or never false-positive (hereafter called "negative"). Person-years at risk were calculated from date of first screen until censoring or end of follow-up. Women contributed person-years at risk to the negative group as long as the screening tests were negative only. Women contributed person-years at risk to the false-positive group from the date of the first false-positive test. Women were censored at death, breast cancer diagnosis, emigration, or end of follow-up on April 17, 2008, whichever came first. Relative risks were tabulated by 1) age at time of false-positive test (50–59 or 60–69 years), 2) type of false-positive test, 3) screen number (first screen, second screen, or third screen or more), and 4) time (years) since the first false-positive test. In the analyses by age at time of false-positive test, type of false-positive test, and screen number, the relevant exposure variable was modeled, for example, for age at time of false-positive test, as $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false-positive at 50–59 years, false-positive at 60–69 years, or negative. For time since false-positive test, the following model was used: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = years after false-positive test (0–2, >2 to 4, >4 to 6, >6 to 8, >8 to 10, >10 to 12, 12 years, or negative).

Second, to take account of a possible temporal trend, separate analyses were made based on screening outcomes in each of the seven approximately biennial invitation rounds [$\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false-positive or negative]. Women with a false-positive test in a previous invitation round were excluded from the analysis of subsequent invitation rounds. Person-years at risk were calculated from date of screening in the respective invitation round until censoring or end of follow-up. Women contributed person-years at risk to the negative group if they screened negative and to the false-positive group if the screening test turned out to be false positive. Women were censored at death, breast cancer diagnosis, emigration, or end of follow-up on April 17, 2008, whichever came first.

Third, to take account of the short follow-up time after the last two invitation rounds, we analyzed two groups of screened women balanced in follow-up time, that is, women screened during the period January 1, 1994, to December 31, 1998, and followed up to December 31, 2000, and women screened during the period January 1, 2001, to December 31, 2005, and followed up to December 31, 2007 (model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false positive or negative). Women with a false-positive test during the period 1994–1998 were excluded from the analysis of data from 2001 to 2005.

Regression analyses were performed in SAS procedure PROC GENMOD. The χ^2 test was used to compare differences in tumor size, receptor, and nodal status between the breast cancers diagnosed in the two groups of women. The statistical calculations were done using SAS version 9.1 (SAS Institute Inc, Cary, NC).

Results

A total of 58 003 women were included in the analysis, and from the time of their first screen, they accumulated 631 039 person-years at risk. Women with negative tests contributed 580 450 person-years at risk, and women with false-positive tests contributed 50 589 person-years at risk. The total number of cancers in women with negative tests was 1969, giving an absolute cancer rate of 339/100 000 person-years at risk, whereas the number for women with false-positive tests was 295, giving an absolute cancer rate of 583/100 000 person-years at risk (Table 1). The mean follow-up time was 10.9 years (10.7 for false positive and 10.9 for negative). The recall rate decreased from 6.78% to 2.26% in the study period, the detection rate varied between 1.20% and 0.58%, and the false-positive rate decreased from 5.58% to 1.39% (Figure 1).

The relative risk of breast cancer adjusted for age at diagnosis after any type of false-positive test was higher than for women with negative tests (RR = 1.67, 95% CI = 1.45 to 1.88, Table 1). The relative risk of breast cancer for women having a false-positive test at age 50–59 and 60–69 years was higher than for women testing negative in those age categories (50–59 years: RR = 1.65, 95% CI = 1.40 to 1.95; 60–69 years: RR = 1.69, 95% CI = 1.42 to 2.01). The relative risk of breast cancer was 1.69 after a type 1 false-positive test and 1.47 after a type 2 false-positive test, compared with women who tested negative. Women with a false-positive test at first screen had a lower relative risk than women who tested false-positive at later screens (first screen: RR = 1.34, 95% CI = 1.10 to 1.63; second screen: RR = 1.87, 95% CI = 1.49 to 2.36; third screen or more: RR = 2.00, 95% CI = 1.65 to 2.42). In the 2 years following a false-positive test, there was no statistically significant difference in breast cancer incidence between women with false-positive and those with negative screening tests (RR = 1.06, 95% CI = 0.73 to 1.52). The relative risk statistically significantly increased 2–4 years after a false-positive test (RR = 2.29, 95% CI = 1.79 to 2.93, $P < .001$), whereas the increase was not statistically significant after 4–6 years (RR = 1.28, 95% CI = 0.92 to 1.78, $P = .11$). During the long-term follow-up 6 or more years after a false-positive test, the risk of breast cancer was statistically significantly increased, with relative risk estimates varying from 1.58 to 2.30. Women testing false positive at first screen had an increased risk of breast cancer detection at next screen (RR =1.36, 95% CI = 0.80 to 2.36), and this risk tended to increase with increasing screen number for the false-positive test, although this result is based on very small numbers (data not shown).

The age-adjusted relative risk of breast cancer for women with a false-positive test in the first invitation round was moderately higher than that for women who screened negative in this round (RR = 1.38, 95% CI = 1.14 to 1.68, $P = .001$), reflecting that all screens in this invitation round were initial screens. Somewhat higher relative risks of breast cancer with false-positive tests compared with negative tests were found for women screened in the next invitation rounds (second, third, fourth, fifth round: RR = 1.78, 2.06, 1.83, and 1.68, respectively). In these middle invitation

Table 1. Relative risk of breast cancer (invasive and DCIS) for women with false-positive screening tests vs women with negative screening tests*

Cohort	Person-years at risk	Breast cancer total	RR, Crude	Adjusted RR (95% CI)	P
First analysis†					
Negative test	580 450	1969	1.0	1.0 (Referent)	
False positive test	50 589	295	1.72	1.67 (1.45 to 1.88)	<.001
By age at false-positive test, y‡					
50–59	28 327	154	1.60	1.65 (1.40 to 1.95)	<.001
60–69	22 261	141	1.86	1.69 (1.42 to 2.01)	<.001
By type of false-positive test§					
Type 1	45 550	269	1.74	1.69 (1.49 to 1.92)	<.001
Type 2	5038	26	1.52	1.47 (1.00 to 2.16)	.051
By screen number at false-positive test‖					
First screen	21 944	108	1.45	1.34 (1.10 to 1.63)	.003
Second screen	11 544	76	1.94	1.87 (1.49 to 2.36)	<.001
Third or more screen	17 101	111	1.91	2.00 (1.65 to 2.42)	<.001
By time since false-positive test, y¶					
0 to 2	9375	30	0.94	1.06 (0.73 to 1.52)	.75
>2 to 4	8839	66	2.20	2.29 (1.79 to 2.93)	<.001
>4 to 6	8101	36	1.31	1.28 (0.92 to 1.78)	.11
>6 to 8	7189	41	1.68	1.58 (1.16 to 2.16)	.001
>8 to 10	6155	40	1.92	1.73 (1.26 to 2.36)	<.001
>10 to 12	4979	44	2.61	2.30 (1.71 to 3.11)	<.001
>12	5951	38	1.88	1.64 (1.19 to 2.28)	.001
Second analysis#					
Invitation round					
First: April 1, 1991 to April 23, 1993					
Negative test	361 106	1279	1.0	1.0 (Referent)	
False-positive test	21 944	108	1.39	1.38 (1.14 to 1.68)	.001
Second: April 26, 1993 to May 30, 1995					
Negative test	277 236	1046	1.0	1.0 (Referent)	
False-positive test	25 400	148	1.54	1.78 (1.41 to 2.25)	<.001
Third: June 1, 1995 to March 24, 1997					
Negative test	236 083	922	1.0	1.0 (Referent)	
False-positive test	23 392	152	1.66	2.06 (1.54 to 2.77)	<.001
Fourth: March 25, 1997 to April 19, 1999					
Negative test	202 423	805	1.0	1.0 (Referent)	
False-positive test	20 661	141	1.72	1.83 (1.30 to 2.58)	.0006
Fifth: April 20, 1999 to March 31, 2001					
Negative test	162 636	659	1.0	1.0 (Referent)	
False-positive test	16 532	122	1.82	1.68 (1.06 to 2.65	.027
Sixth: April 1, 2001 to May 31, 2003					
Negative test	127 072	535	1.0	1.0 (Referent)	
False-positive test	12 021	88	1.74	1.23 (0.64 to 2.38)	.53
Seventh: June 1, 2003 to December 31, 2005					
Negative test	76 295	363	1.0	1.0 (Referent)	
False-positive test	6595	50	1.59	0.48 (0.11 to 1.91)	.29
Third analysis**					
Technology period					
January 1, 1994 to December 31, 1998, follow-up ends December 31, 2000					
Negative test(s)	157 542	470	1.0	1.0 (Referent)	
False-positive test(s)	9604	46	1.61	1.65 (1.22 to 2.24)	.001
January 1, 2001 to December 31, 2005, follow-up ends December 31, 2007					
Negative test(s)	160 916	555	1.0	1.0 (Referent)	
False-positive test(s)	5486	23	1.22	1.31 (0.87 to 2.00)	.2

* Adjusted by attained age, Copenhagen, Denmark, 1991–2008. All statistical tests were two-sided. DCIS = ductal carcinoma in situ; RR = relative risk.

† Poisson regression for the entire dataset. Ln(breast cancer incidence) = $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where a is age at time of diagnosis in 5-year age groups (50–54, 55–59, 60–64 years, or 65–69, 70–74, 75–79, 80–84, 85–89 years), and s is exposure status (negative test or positive test).

‡ Model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false-positive at age 50–59 years, false-positive at age 60–69 years, or negative.

§ Model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = type 1 false-positive, type 2 false-positive or negative. Type 1 false-positive test = women cleared of suspicion of breast cancer at the triple test (First assessment); Type 2 false-positive test = women cleared of suspicion at surgical biopsy (Second assessment).

‖ Model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false positive at first screen, false positive at second screen, false positive at third screen or more, or negative.

¶ Model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = 0–2 years after false-positive test, 2–4 years after false-positive test, 4–6 years after false-positive test, 6–8 years after false-positive test, 8–10 years after false-positive test, 10–12 years after false-positive test, 12 years or more after false-positive test, or negative.

\# Poisson regression for each invitation round separately. Women with a false-positive test in a given invitation round were excluded from subsequent invitation rounds in this analysis. Model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false-positive or negative.

** Poisson regression for women screened 1994–1998 and followed up to end of 2000, and for women screened in 2001–2005 and followed up to end of 2007. Women with a false-positive test in 1994–1998 were excluded from the analysis of data from 2001 to 2005. Model: $\ln(\lambda_{as}) = \alpha + \beta_a a + \beta_s s$, where s = false positive or negative.
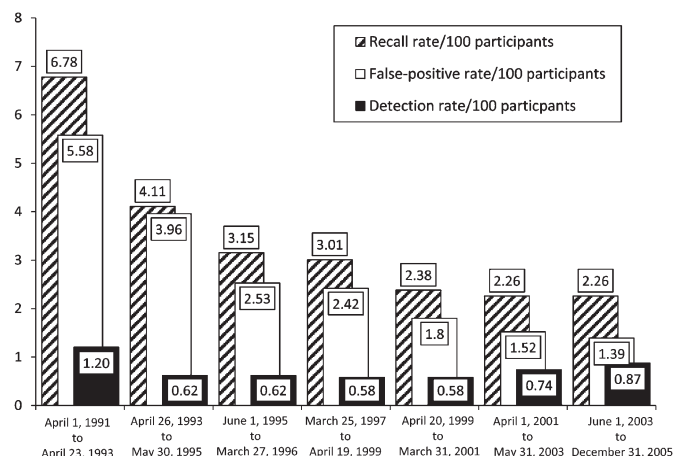
Figure 1. Recall rate, false-positive rate, and cancer detection rate per 100 women screened by invitation round. Copenhagen, Denmark, 1991–2005. Recall rate = number of women recalled for further assessment/number of women screened; false-positive rate = number of women with false-positive test/number of women screened; cancer detection rate = number of women with screen detected cancer/number of women screened.

rounds, the majority of the screens were subsequent, not first, screens. During the last invitation rounds in the series, the relative risks were compatible with unity (sixth round: RR = 1.23, 95% CI = 0.64 to 2.38, P = .53; seventh round: RR = 0.48, 95% CI = 0.11 to 1.91, P = .29), which may reflect a true temporal trend because new screening technology was introduced at this time, but it may also reflect the fact that the follow-up periods were short for women screened in these invitation rounds. With the introduction of new, and better, screening technology, detection of true positives among the assessed may be increased, hence, also increasing the proportion of true negatives in the false-positive population. However, with the short follow-up and wide confidence intervals, the result for the seventh round cannot be assumed to indicate an increased protection among women with false-positive test compared with women with negative test.

We therefore analyzed separately, women screened between January 1, 1994, and December 31, 1998, and followed up until December 31, 2000, and women screened between January 1, 2001, and December 31, 2005, and followed up until December 31, 2007. For the first period, women with a false-positive test had a statistically significantly higher age-adjusted relative risk than women with a negative test (RR = 1.65, 95% CI = 1.22 to 2.24, P = .001), whereas in the second period, the relative risk of breast cancer for women with a false-positive test was not statistically significantly different from that for women with a negative test (RR = 1.31, 95% CI = 0.87 to 2.00).

There were no statistically significant differences in tumor size, receptor, or nodal status between invasive breast cancers in women with false-positive and in women with negative screening tests (Table 2).

Discussion

Across the full follow-up period, women who had experienced a false-positive screening test had a 67% higher risk of breast cancer than women without a false-positive screening test. This excess

risk did not derive from interval cancers, that is, cancers diagnosed within 2 years of the false-positive test. There was, however, a marked excess risk 2–4 years later, for which the comparison with previous mammograms could reveal eventual changes in the suspected tissue. The interval cancer risk after 4–6 years was not statistically significantly increased either. The most remarkable finding in this study was the statistically significantly increased risk of breast cancer in women 6 years or more after their false-positive screening test. If the false-positive test occurred at the first screen, the relative risk of breast cancer was lower than if the false-positive test occurred at a later screen, which might be explained by the opportunity to compare later with earlier mammograms at subsequent screens, leading to identification of more true positives at subsequent screens. Thus, at the initial screen, women with false-positive tests might be a less selected population, that is, containing more true-negative tests, than at subsequent screens.

It was encouraging that the 65% excess risk of breast cancer in women with false-positive tests in the late 1990s (1994–1998) dropped to a 31% non-statistically significant excess risk in the early 2000s (2001–2005). Over the same periods, the detection rate increased and the false-positive rate decreased. These important improvements coincided with major changes in the screening technology, that is, the introduction of high-frequency ultrasound devices in 2001, stereotactic breast biopsy in 2002, and bi-directional mammography as standard in 2004 (11).

The existence of complete data from the mammography program as well as population-based cancer data enabled us to map exposure and outcome on an individual level, with no loss to follow-up. However, this study also had some limitations. Tumor size and receptor status were missing for approximately 5% of cancer patients, but in view of the even distribution between tumors detected in women with false-positive tests and tumors detected in women with negative tests, this was not considered to be of importance. To get two equal periods of comparison for the two periods of technology assessment, the follow-up after screening had to be reduced to 2 years, which resulted in smaller sample sizes and wider confidence intervals. Therefore, similarity between the two periods in increased risk of breast cancer for women with false-positive tests could not be ruled out (P = .15). A comparative analysis of incidence of breast cancer by calendar period of the false-positive test would have been useful, but the comparison would have been hampered by the shorter follow-up time for the women with a false-positive test in later years compared with women with a false-positive test in earlier years. Unfortunately, we could not tabulate false-positive tests by left or right breast because these data were not part of the screening mammography dataset.

In 1988, Peeters et al. (8) reported a relative risk of breast cancer of 2.72 (P = .0006) for women with a false-positive test. They compared 462 women with false-positive tests with a reference group of 1865 true-negative women, with a mean follow-up time of 5 years. Peeters et al. (8) argued that in 12 of the 16 cancers in women with false-positive tests, the malignancy had been present already at the time of the initial referral, which was not shown in this study, in which there was no increase in the relative risk in the first 2 years after the false-positive test. There was, however, an increased risk at the second screen, and those instances might have included cancers present at the time of the initial assessment.

Table 2. Invasive breast cancers by screening status, tumor size, receptor, and nodal status*

Screening and tumor characteristics	False-positive screening tests, No. (%)	Negative screening tests, No. (%)	Total, No. (%)	P
Total No. of cancers	274 (100)	1817 (100)	2091 (100)	.95
Data available from DBCG	258 (94.2)	1720 (94.7)	1978 (94.6)	
Tumor size				
≤ 10 mm	67 (26.0)	438 (25.4)	505 (25.5)	.75
10–20 mm	113 (43.8)	756 (44.0)	869 (43.9)	
>20 mm	70 (27.1)	450 (26.2)	520 (26.3)	
Neoadjuvant	1 (0.4)	3 (0.2)	4 (0.2)	
NA/missing	7 (2.7)	73 (4.2)	80 (4.0)	
ER status				
Negative	35 (13.6)	305 (17.7)	340 (17.2)	.15
Positive	214 (83.0)	1335 (77.6)	1549 (78.3)	
NA/missing	9 (3.5)	80 (4.7)	89 (4.5)	
PgR status				
Negative	75 (29.1)	504 (29.3)	579 (29.3)	.85
Positive	107 (41.5)	684 (39.8)	791 (40.0)	
NA/missing	76 (29.5)	532 (30.9)	608 (30.7)	
Nodal status				
Negative	158 (61.2)	1035 (60.2)	1193 (60.3)	.54
Positive	90 (34.9)	589 (34.2)	679 (34.3)	
NA/missing	10 (3.9)	96 (5.6)	106 (5.4)	
HER2				
Negative	62 (24.0)	397 (23.1)	459 (23.2)	.91
Positive	15 (5.8)	109 (6.3)	124 (6.3)	
NA/missing	181 (70.2)	1214 (70.6)	1395 (70.5)	

* Copenhagen, Denmark, 1991–2008. DBCG = Danish Breast Cancer Cooperative Group; ER = Estrogen receptor; NA = not applicable; PgR = Progesterone receptor. Statistical comparison of data between the two groups (false-positive test and negative test) was done with a χ^2 test. All statistical tests were two-sided.

However, in this study, only 40% of the breast cancers in women with false-positive tests occurred within 5 years of follow-up.

In contrast to this study, McCann et al. (10) reported a higher risk estimate for interval cancer (OR = 3.19, 95% CI = 2.34 to 4.35) in women with false-positive tests at the first screen than at the second screen (OR = 2.15, 95% CI = 1.55 to 2.98). The McCann et al. (10) data were exclusively based on screens from the prevalence round and hence included only women with false-positive tests at their initial screen. The screening interval in East Anglia (15) was approximately 3.5 years compared with only 2 years in Denmark, and the East Anglia program operated with early recall, typically 6 or 12 months after assessment (15), which has not been practiced in Denmark. These differences in the organization of the programs may explain the differences in the findings concerning the interval cancers. Barlow et al. (16) found a statistically significantly increased risk (OR = 1.69, 95% CI = 1.47 to 1.94) for breast cancer within 1 year from a negative screen among women with a false-positive test at the screen previous to the negative screen, which is consistent with the results of this study.

Groenendijk et al. (9) did not find any difference in risk of breast cancer between women with a false-positive test and women with a negative test. The study population was based on 188 women, and the women were defined as false positives after having had an excision (54%), that is, a type 2 false positive, or having had further bi-directional mammography and/or ultrasound (46%), that is, a type 1 false positive.

The long-term excess breast cancer risk of 67% found in this study was from a population-based screening mammography program for which, in an earlier study, the cumulative risk of a false-positive test was found to be close to 16%, assuming independence between the screens. The hypotheses of independence between the outcomes of subsequent screens was tested and published earlier (17). In general, however, the proportion of screened women with false-positive tests does vary considerably across screening settings.

Elmore et al. (18), followed by Christiansen et al. (19), reporting on a large health maintenance organization in New England, found that a false-positive test occurred in 6.5% of mammograms and that the cumulative risk of a false-positive test over 10 annual screens was 49.1% (95% CI = 40.3% to 64.1%). Hubbard et al. (1) collected data from seven mammography registers in the United States and, on the basis of different models, concluded that the cumulative risk of a false-positive mammogram over 10 screens varied between 58% and 77%, with an estimate of 63% from the model with the most reasonable assumptions. This can be compared with data from the population-based breast screening program in Barcelona, Spain (20), in which the cumulative risk of a false-positive test result was 32% over 10 biennial screens, and with data from the Norwegian Breast Screening Program, for which Hofvind et al. (21) reported a cumulative risk of 21% over 10 biennial screens. A cumulative risk of 6% can be estimated from data reported in a Netherlands study (22). Considering the large difference in recall rate between the United States and Denmark, one may, on the basis of who is selected for recall, expect the breast cancer risk in the large group of US women with a false-positive test to be closer to that of test-negative women than what we found for the more restricted group of Copenhagen women with a false-positive test. Nevertheless, because women's biology and the

screening technology are fairly similar in the two countries, the large US population of women with false-positive tests is expected to be heterogeneous and therefore to include certain subgroups of women with excess risks similar to those found for the Copenhagen women.

The excess breast cancer risk in women with false-positive tests may be attributable to misclassification of malignancies already present at the baseline assessment, as indicated in the short-term Netherlands study (8), or to a biological susceptibility for developing breast cancer in some women without malignancies at baseline. Earlier results (23–25) show, for example, both an increased risk of false-positive tests and breast cancer among hormone users. In this study, the finding of a more than doubled risk at the first screen following the false-positive test favors the hypothesis of misclassification, as does the fact that the excess risk was higher in the early technology phase than in the late technology phase, in which high-resolution ultrasound and stereotactic biopsies were available. Misclassification should of course relate to the same breast that later contracts breast cancer, but, unfortunately, we did not have access to mammography data on location of the initial suspicious finding. The persistent excess breast cancer risk up to 12 years and more after the baseline false-positive test favors the hypothesis of biological susceptibility. This is consistent with the excess breast cancer risk found for women with benign breast lesions (2–7).

The experience of a false-positive test causes anxiety (26–29), which may discourage women from attending screening regularly. The long-term excess risk of breast cancer found in women with false-positive tests stresses the need for their adherence to regular screening. In Copenhagen, women with false-positive tests were informed that no malignancy was found and they were re-invited to the next screening round, just as women with negative screening tests were. During the first five biennial rounds of the Copenhagen program, women with false-positive tests attended their next screening round to exactly the same extent as did women with negative screening tests (30).

In conclusion, based on the findings in this study, it may be beneficial to actively encourage women with false-positive tests to continue to attend regular screening. This consideration has to be weighed against the risk of causing extra anxiety. Furthermore, it is important to collect longer follow-up data for women screened after the introduction of a new screening technology.

References

1. Hubbard RA, Miglioretti DL, Smith RA. Modelling the cumulative risk of a false-positive screening test. *Stat Methods Med Res.* 2010;19(5):429–449.
2. Potter JF, Slimbaugh WP, Woodward SC. Can breast carcinoma be anticipated? A follow-up of benign breast biopsies. *Ann Surg.* 1968;167(6):829–838.
3. Hutchinson WB, Thomas DB, Hamlin WB, Roth GJ, Peterson AV, Williams B. Risk of breast cancer in women with benign breast disease. *J Natl Cancer Inst.* 1980;65(1):13–20.
4. Roberts MM, Jones V, Elton RA, Fortt RW, Williams S, Gravelle IH. Risk of breast cancer in women with history of benign disease of the breast. *Br Med J (Clin Res Ed).* 1984;288(6413):275–278.
5. Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med.* 2005;353(3):229–237.
6. Dorjgochoo T, Deming SL, Gao YT, et al. History of benign breast disease and risk of breast cancer among women in China: a case-control study. *Cancer Causes Control.* 2008;19(8):819–828.
7. Kabat GC, Jones JG, Olson N, et al. A multi-center prospective cohort study of benign breast disease and risk of subsequent breast cancer. *Cancer Causes Control.* 2010;21(6):821–828.
8. Peeters PH, Mravunac M, Hendriks JH, Verbeek AL, Holland R, Vooijs PG. Breast cancer risk for women with a false positive screening test. *Br J Cancer.* 1988;58(2):211–212.
9. Groenendijk RP, Kochen MP, van Engelenburg KC, et al. Detection of breast cancer after biopsy for false-positive screening mammography. An increased risk? *Eur J Surg Oncol.* 2001;27(1):17–20.
10. McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res.* 2002;4(5):R11.
11. Utzon-Frank N, Vejborg I, von Euler-Chelpin M, Lynge E. Balancing sensitivity and specificity: sixteen year's of experience from the mammography screening programme in Copenhagen, Denmark. *Cancer Epidemiol.* 2011;35(5):393–398.
12. Breast Cancer Screening Consortium. BCSC Glossary of Terms. http://breastscreening.cancer.gov/data/bcsc_data_definitions.pdf. Accessed November 23, 2011.
13. European Commission. *European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis.* 4th ed. Luxembourg: European Communities; 2006.
14. The American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS) Atlas. http://www.acr.org/SecondaryMainMenuCategories/ACRStore/FeaturedCategories/QualityandSafety/birads_atlas/BIRADSFAQs.aspx. Accessed July 11, 2011.
15. Ong GJ, Austoker J, Michell M. Early rescreen/recall in the UK National Health Service breast screening programme: epidemiological data. *J Med Screen.* 1998;5(3):146–155.
16. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst.* 2006;98(17):1204–1214.
17. Njor SH, Olsen AH, Schwartz W, Vejborg I, Lynge E. Predicting the risk of a false-positive test for women following a mammography screening programme. *J Med Screen.* 2007;14(2):94–97.
18. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med.* 1998;338(16):1089–1096.
19. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. *J Natl Cancer Inst.* 2000;92(20):1657–1666.
20. Castells X, Molins E, Macia F. Cumulative false positive recall rate and association with participant related factors in a population based breast cancer screening programme. *J Epidemiol Community Health.* 2006;60(4):316–321.
21. Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. *Cancer.* 2004;101(7):1501–1507.
22. National Evaluation Team for Breast Cancer Screening (NETB). *National Evaluation of Breast Cancer Screening in the Netherlands 1990-2007.* Rotterdam, the Netherlands: Erasmus MC; 2009.
23. Glass AG, Lacey JV Jr., Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J Natl Cancer Inst.* 2007;99(15):1152–1161.
24. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA.* 2002;288(3):321–333.
25. Njor SH, Pedersen AT, Schwartz W, Hallas J, Lynge E. Minimizing misclassification of hormone users at mammography screening. *Int J Cancer.* 2009;124(9):2159–2165.
26. Aro AR, Pilvikki AS, van Elderen TM, van der Ploeg E, van der Kamp LJ. False-positive findings in mammography screening induces short-term distress—breast cancer-specific concern prevails longer. *Eur J Cancer.* 2000;36(9):1089–1097.
27. Barton MB, Morley DS, Moore S, et al. Decreasing women's anxieties after abnormal mammograms: a controlled trial. *J Natl Cancer Inst.* 2004;96(7):529–538.

28. Salz T, Richman AR, Brewer NT. Meta-analyses of the effect of false-positive mammograms on generic and specific psychosocial outcomes. *Psychooncology*. 2010;19(10):1026–1034.

29. van der Steeg AF, Keyzer-Dekker CM, De VJ, Roukema JA. Effect of abnormal screening mammogram on quality of life. *Br J Surg*. 2011;98(4):537–542.

30. Andersen SB, Vejborg I, von Euler-Chelpin M. Participation behaviour following a false positive test in the Copenhagen mammography screening programme. *Acta Oncol*. 2008;47(4):550–555.

Notes

BLT is currently employed by NovoNordisk A/S.

Affiliations of authors: Department of Public Health, University of Copenhagen, Copenhagen K, Denmark (MvE-C, LMR, BLT); Centre of Diagnostic Imaging, Rigshospitalet, University Hospital Copenhagen, Copenhagen Ø, Denmark (IV).